Mathematische und statistische Methoden I

Dr. Malte Persike

persike@uni-mainz.de
lordsofthebortz.de
twitter.com/methodenlehre
tinyurl.com/gplusmethodenlehre

Sprechstunde jederzeit nach Vereinbarung und nach der Vorlesung

Wallstr. 3, 6. Stock,
Raum 06-206

WiSe 2011/2012

Fachbereich Sozialwissenschaften
Psychologisches Institut
Johannes Gutenberg Universität Mainz

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Ordinalskala

Definition

- Bei einer Ordinalskala können die Realisationen einer Variablen (natürlich) **geordnet** werden

- Die Zuordnung der Zahlen zu den Ausprägungen spiegelt die Ordnung wieder Abstände zwischen den Zahlen können **nicht interpretiert** werden

- Die Anwendung von Rechenoperationen auf die Werte einer ordinalskalierten Variablen ist unter bestimmten Voraussetzungen erlaubt, aber im Allgemeinen eher wenig sinnvoll

Ordinalskala

Beispiel

Social Penetration Theory von Altman und Taylor (1958)

(I) *Orientierungsstadium*: Sozial erwünschte Normen und Verhaltensschemata werden ausgetauscht (z.B. Smalltalk)

(II) *Exploratorisch-affektives Stadium*: Partielle Öffnung der eigenen Einstellungs- und Wahrnehmungswelt gegenüber dem Anderen im Hinblick auf private, vor allem aber berufliche und weltanschauliche Inhalte. Weiterhin vorsichtige Prüfung der Interaktionsformen („Bekanntschaftsphase").

(III) *Affektives Stadium*: Intensiver und möglicherweise kritischer Austausch über private und persönliche Themen. Körperliche Zuwendung wie Berühren und Küssen.

(IV) *Stabiles Stadium*: Die Beziehung erreicht ein Plateau, persönliche Inhalte sind geteilt, Verhalten und Emotionen des Anderen vorhersagbar.

(V) *Depenetration*: Zusammenbruch und mögliches Ende der Beziehung, Überwiegen von Kosten gegenüber dem Nutzen.

Ordinalskala

Häufigkeiten

Kennwerte

Grafische Darstellung

Ordinalskala

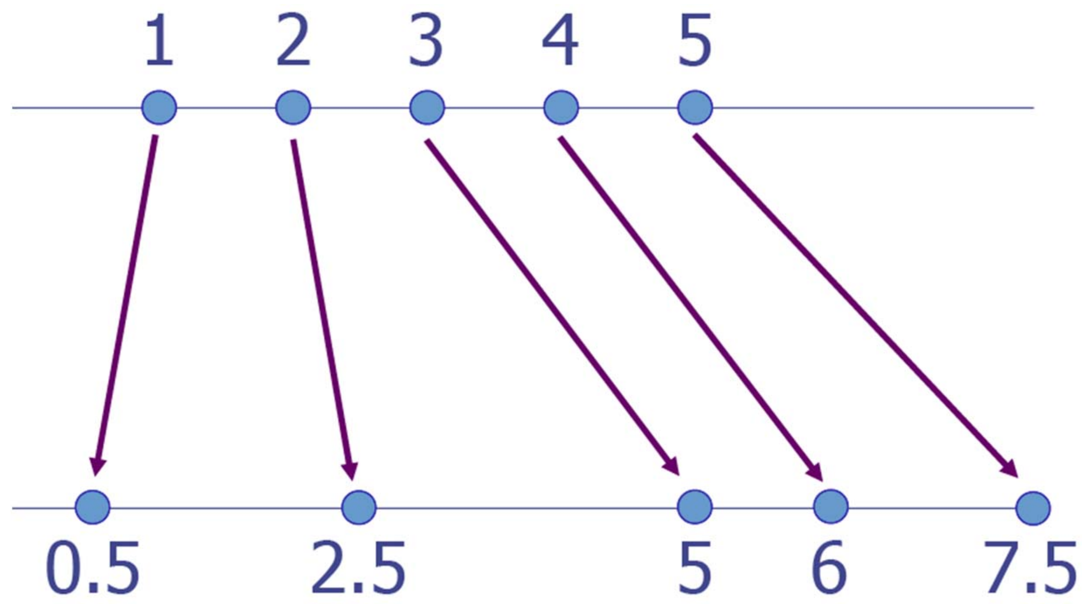
Zulässige Operationen

- Zulässige Operationen sind **Äquivalenzrelationen**, d.h. **Gleich** und **Ungleich**
- **Zudem** erlaubt sind **qualitative Vergleichsrelationen**, d.h. **Größer** oder **Kleiner**
- **Wichtig:** Diese Vergleichsrelationen umfassen **nicht** jede Art quantitativer Vergleiche
- Eine Aussage wie „A ist gleich/ungleich/größer/kleiner B" ist bei einer ordinalskalierten Variable zulässig, nicht aber „A ist viermal so groß wie B".

Ordinalskala

Zulässige Transformationen

Zulässig sind alle **streng monotonen Transformationen**, so dass die Rangordnung der Werte erhalten bleibt.



Ordinalskala

Häufigkeiten

Kennwerte

Grafische Darstellung

Ordinalskala

Kritische Betrachtung

- Bei Ordinalskalen und höheren Skalenniveaus können **Intransitivitäten** auftreten
- **Intransitivität** = Eine angenommene Ordnung gilt nicht für bestimmte einzelne Paarungen
- **Beispiel**: Nahrungskette in chinesischen Restaurants

  Mensch → Hund → Ratte → Mensch (nach Glutamatvergiftung)
- **Lösungen**: Annahme eines niedrigeren Skalenniveaus, Einführung neuer Skalenstufen

Ordinaldaten

Numerische Beschreibung: Häufigkeiten

- Ordinalskalierte Variablen sind sehr häufig diskret und endlich
- Es gelten die bereits eingeführten Notationen und Berechnungsvorschriften für empirische Häufigkeiten
- Neben der Häufigkeitsverteilung kann auch noch die **empirische Verteilungsfunktion** bestimmt werden.
- Diese gibt an, wie viele Beobachtungen kleiner oder gleich einer bestimmten Ausprägung x sind.
- Zur Berechnung der Verteilungsfunktion müssen die Ausprägungen zunächst **der Größe nach geordnet** werden.

Bortz, S. 47

Ordinaldaten

Numerische Beschreibung: Häufigkeiten

- Empirische Häufigkeitsverteilung und Verteilungsfunktion:

Wert von X (geordnet)	$f(X = x_j)$	$F(X \leq x_j)$
x_1	$f(x_1)$	$f(x_1)$
x_2	$f(x_2)$	$f(x_1)+f(x_2)$
...	...	...
x_k	$f(x_k)$	$f(x_1)+f(x_2)+\ldots+f(x_k)$

- Berechnungsvorschrift:
  analog für absolute Vert.funkt. $H(X \leq x_j)$

$$\boxed{F(X \leq x_j) = \sum_{c=1}^{j} f(x_c)}$$

- Für Ordinaldaten gelten die bereits eingeführten Konventionen zur Erstellung von Kreuztabellen

Exkurs: Das Summenzeichen Σ

Notation

- Sei X eine Variable mit k möglichen Ausprägungen
- An n Merkmalsträgern werden nun die Beobachtungen $x_1, x_2, \ldots, x_n$ erhoben
- Die Summe aller n Beobachtungen ist definiert als

$$\boxed{\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n}$$

Exkurs: Das Summenzeichen Σ

Rechenregeln

1. Multiplikation jeder Beobachtung mit einer Konstanten

$$\sum_{i=1}^{n}(a \cdot x_i) = a \cdot x_1 + a \cdot x_2 + \ldots + a \cdot x_n$$
$$= a \cdot (x_1 + x_2 + \ldots + x_n)$$
$$= a \cdot \sum_{i=1}^{n} x_i$$

$$\text{mit } \; a = const.$$

Exkurs: Das Summenzeichen Σ

Rechenregeln

2. Summation einer Konstanten

$$\boxed{\begin{aligned}\sum_{i=1}^{n} b &= \underbrace{b + b + \ldots + b}_{\text{n-mal}} \\ &= n \cdot b\end{aligned}}$$

$$\text{mit b} = const.$$

Exkurs: Das Summenzeichen Σ

Rechenregeln

3. Addition einer Konstanten zu jeder Beobachtung

$$\sum_{i=1}^{n}(x_i+b)=x_1+b+x_2+b+\ldots+x_n+b$$

$$=x_1+x_2+\ldots+x_n+\underbrace{b+b+\ldots+b}_{\text{n-mal}}$$

$$=\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}b=\sum_{i=1}^{n}x_i+n\cdot b$$

$$\text{mit } b=const.$$

Exkurs: Das Summenzeichen Σ

Rechenregeln

4. Addition zweier Variablen

$$\sum_{i=1}^{n}(x_i + y_i) = x_1 + y_1 + x_2 + y_2 + \ldots + x_n + y_n$$

$$= x_1 + x_2 + \ldots + x_n + y_1 + y_2 + \ldots + y_n$$

$$= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

Exkurs: Das Summenzeichen Σ

Rechenregeln

5. Verbindung von Multiplikation und Addition einer Konstanten

$$\sum_{i=1}^{n}(a \cdot x_i + b) = a \cdot \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} b$$

$$= a \cdot \sum_{i=1}^{n} x_i + n \cdot b$$

$$\text{mit } a, b = const.$$

Exkurs: Das Summenzeichen Σ

Geschachtelte Summen - Notation

- Sei X eine Variable mit k möglichen Ausprägungen und Y eine Variable mit m möglichen Ausprägungen
- An n Merkmalsträgern werden die Verbundhäufigkeiten $h(x_i, y_j)$ erhoben, wobei $i=1 \ldots k$ und $j=1 \ldots m$
- Die Summe aller Verbundhäufigkeiten ist

$$\begin{aligned}\sum_{i=1}^{k}\sum_{j=1}^{m} h(x_i, y_j) = \quad & h(x_1, y_1) + h(x_1, y_2) + \ldots + h(x_1, y_m) + \\ & h(x_2, y_1) + h(x_2, y_2) + \ldots + h(x_2, y_m) + \\ & \ldots + \\ & h(x_k, y_1) + h(x_k, y_2) + \ldots + h(x_k, y_m)\end{aligned}$$

Exkurs: Das Summenzeichen Σ

Geschachtelte Summen - Rechenregeln

1. Kommutativgesetz

$$\boxed{\sum_{i=1}^{k}\sum_{j=1}^{m} h\left(x_i, y_j\right) = \sum_{j=1}^{m}\sum_{i=1}^{k} h\left(x_i, y_j\right)}$$

2. Keine Trennung von geschachtelten Summen

$$\boxed{\sum_{i=1}^{k}\sum_{j=1}^{m} h\left(x_i, y_j\right) \neq \underbrace{\sum_{i=1}^{k} h\left(x_i, y_j\right) + \sum_{j=1}^{m} h\left(x_i, y_j\right)}_{\text{nicht definiert}}}$$

Häufigkeiten

Kreuztabellen

Kennwerte

Grafische Darstellung

Exkurs: Das Summenzeichen Σ

Rechnen mit Häufigkeiten (am bivariaten Beispiel)

Anzahl Beobachtungen:

$$n = h(\bullet,\bullet) = \sum_{i=1}^{k_x}\sum_{j=1}^{k_y} h(x_i, y_j)$$

Randhäufigkeiten für x:
analog für $f(x_i, \bullet)$

$$h(x_i,\bullet) = \sum_{j=1}^{k_y} h(x_i, y_j)$$

Randhäufigkeiten für y:
analog für $f(y_j, \bullet)$

$$h(\bullet, y_j) = \sum_{i=1}^{k_x} h(x_i, y_j)$$

Darüber hinaus gilt:

$$\sum_{i=1}^{k_x}\sum_{j=1}^{k_y} f(x_i, y_j) = 1$$

Ordinalskala
Häufigkeiten
Kennwerte
Grafische Darstellung

Ordinaldaten

Numerische Beschreibung: Kennwerte

- Maße der zentralen Tendenz
  - Median
- Andere Lagemaße
  - Modalwert
  - Extrema (Minimum, Maximum)
  - Quantile, Quantilsrang
- Streuungsmaße (Dispersionsmaße)
  - Spannweite
  - (Halber) Interquartilsabstand

Bortz, S. 26 – 28

Ordinaldaten

Numerische Beschreibung: Median

- Mindestens 50% der Beobachtungen einer Variablen sind kleiner oder gleich dem Median
- Mindestens 50% der Beobachtungen einer Variablen sind größer oder gleich dem Median
- Notation: x_{med} oder $\tilde{x}$
- **Problem**: Bei einer geraden Zahl von Beobachtungen ist der Median nicht eindeutig. Er liegt dann zwischen den Beobachtungen

Ordinalskala

Häufigkeiten

Kennwerte

Grafische Darstellung

Ordinaldaten

Numerische Beschreibung: Median

- Der Median stimmt häufig mit keiner beobachteten Ausprägung überein
- Median (und auch der Modalwert) sind äquivariant gegenüber gewissen (z.B. linearen) Transformationen
- Insbesondere
  1. Addition einer Konstanten a zu allen n Messwerten $x_1 \dots x_n$

     $$\widetilde{x+a} = \tilde{x} + a$$

  2. Multiplikation aller n Messwerte $x_1 \dots x_n$ mit einer Konstanten a

     $$\widetilde{x \cdot a} = \tilde{x} \cdot a$$

Bortz, S. 33 – 34

Ordinaldaten

Numerische Beschreibung: Quantile

- Quantile sind Zahlen, die einen Datensatz mit n Beobachtungen in bestimmtem Verhältnis teilen
- p-Quantil $(0 \leq p \leq 1)$ besitzt folgende Eigenschaften:
  1. Mindestens $n{\cdot}p$ Beobachtungen sind kleiner oder gleich dem Quantil
  2. Mindestens $n{\cdot}(1-p)$ Beobachtungen sind größer oder gleich dem Quantil
- Notation: x_p (z. B. $x_{0.75}$)
- Je nach der Anzahl von Unterteilungen unterscheidet man **Percentile** (100er Einteilung), **Dezentile** (10er Einteilung) und **Quartile** (4er Einteilung)

Ordinalskala

Häufigkeiten

Kennwerte

Grafische Darstellung

Ordinaldaten

Numerische Beschreibung: Quantile

Wichtige Quantile sind:

- Perzentile: $x_{.01}, x_{.02}, \ldots, x_{.99}, x_{1.0}$
  bzw. $x_{1\%}, x_{2\%}, \ldots, x_{99\%}, x_{100\%}$
- Minimum (0. Quartil, 0% Perzentil) und Maximum (4. Quartil)
- Median (2. Quartil, 50% Perzentil)
- 25% Perzentil (1. Quartil, unteres Quartil) und 75% Perzentil (3. Quartil, oberes Quartil)
- Dezile: $x_{.10}, x_{.20}, \ldots, x_{.90}$

Ordinaldaten

Quantile – A cautionary note about conventions

In Literatur und Softwarepaketen sind die Berechnungs-vorschriften für Quantile häufig **unterschiedlich definiert** oder sogar fehlerhaft.

Maß	Bortz	Excel	SPSS
Median	137.5	137.5	137.5
1. Quartil	130.5	132.25	128.75
3. Quartil	146	145	147

Für einen Beispieldatensatz mit n=12.

Ordinaldaten

Der Quantilsrang

- Als ***p*-Quantil** war diejenige Ausprägung x der Variablen X definiert, die die Daten in einen Anteil von p Datenwerten unterhalb oder gleich der Ausprägung x sowie $1\text{-}p$ Datenwerten oberhalb oder gleich x teilt.
- Besonders bei angewandten Fragestellungen ist oft auch die **entgegengesetzte Sichtweise** relevant.
- **Beispiel:** Eine Person habe in einem Leistungstest einen Wert von 105 Punkten erzielt. Wie viele Personen in der Stichprobe sind nun besser/schlechter?
- Dies kann über den **Quantilsrang** ermittelt werden.

Ordinaldaten

Der Quantilsrang – Verfahren der Rangbildung

- Bei der Rangbildung von k Ausprägungen $x_1 \ldots x_k$ einer Variablen X können maximal k Ränge vergeben werden.
- Per Konvention erhält die numerisch niedrigste Ausprägung von X den Rangplatz 1, die höchste den Rangplatz k (**kleinere Zahl = kleinerer Rang**).
- Bei mehreren gleichen Werten („**Ties**“) von X wird der mittlere Rangplatz vergeben nach der Regel:

  Es gebe m gleiche Werte von X. Wären sie unterschiedlich und direkt aufeinander folgend, erhielten sie die Rangplätze $rg_j \ldots rg_{j+m-1}$. Der **mittlere Rang** ist dann

$$\boxed{rg_{Tie} = \frac{1}{m} \sum_{i=rg_j}^{rg_j+m-1} rg_i}$$

Ordinaldaten

Der Quantilsrang – Berechung

- Nach der Berechnung der Rangzahl $rg(x)$ eines Merkmalsträgers ermittelt man seinen Quantilsrang p über

$$\boxed{p = \frac{rg(x) - 0.5}{n}}$$

n = Stichprobengröße (d.h. der maximale Rang)

- **Problem:** p reicht nicht von 0 bis 1, sondern liegt in einem etwas schmaleren Bereich, abhängig von der Größe des n.

- Die **Korrekturformel** für den Quantilsrang behebt dieses Problem

$$\boxed{p_{corr} = \frac{p - p_{\min}}{p_{\max} - p_{\min}}}$$

Bortz, S. 32

Ordinaldaten

Numerische Beschreibung: Spannweite

- Die **Spannweite** d_k ist die Differenz zwischen dem kleinsten und größten Wert aller Ausprägungen.
- Sie ist definiert als:

$$\boxed{d_k = x_{\max} - x_{\min}}$$

- Die Spannweite ist nicht identisch mit der Anzahl unterschiedlicher Ausprägungen.
- Die Spannweite ist eher uninformativ, da sie nur zwei von k Ausprägungen berücksichtigt.

Ordinaldaten

Numerische Beschreibung: Interquartilsabstand

- Der **Interquartilsabstand** d_q ist die Differenz zwischen dem 1. und 3. Quartil

- Er ist definiert als

$$\boxed{d_q = x_{.75} - x_{.25}}$$

- Manchmal wird auch ein **halber Interquartilsabstand** berechnet als $d_q/2$.

Ordinaldaten

Grafische Beschreibung: Empirische Verteilungsfunktion

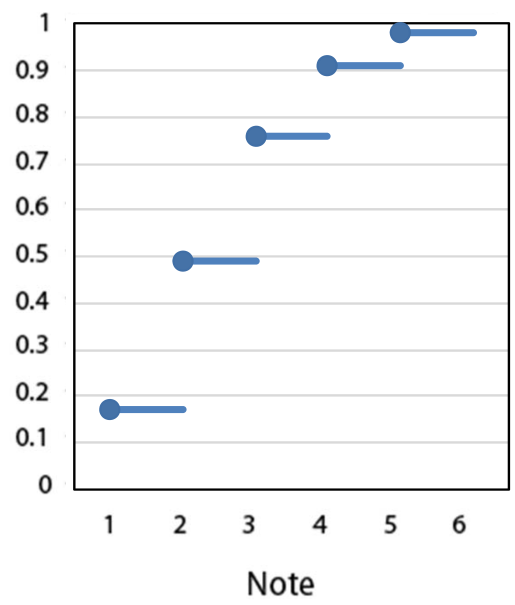
Die empirische Verteilungsfunktion bei k Realisationen ist

$$F(X \leq x_j) = F(x_j) = \sum_{c=1}^{j} f(x_c)$$

mit $j = 1 \ldots k$

Zur grafischen Darstellung werden also die empirischen relativen Häufigkeiten aufsummiert

Note x	h(x)	f(x)
1	7	0.17
2	13	0.32
3	11	0.27
4	6	0.15
5	3	0.07
6	1	0.02



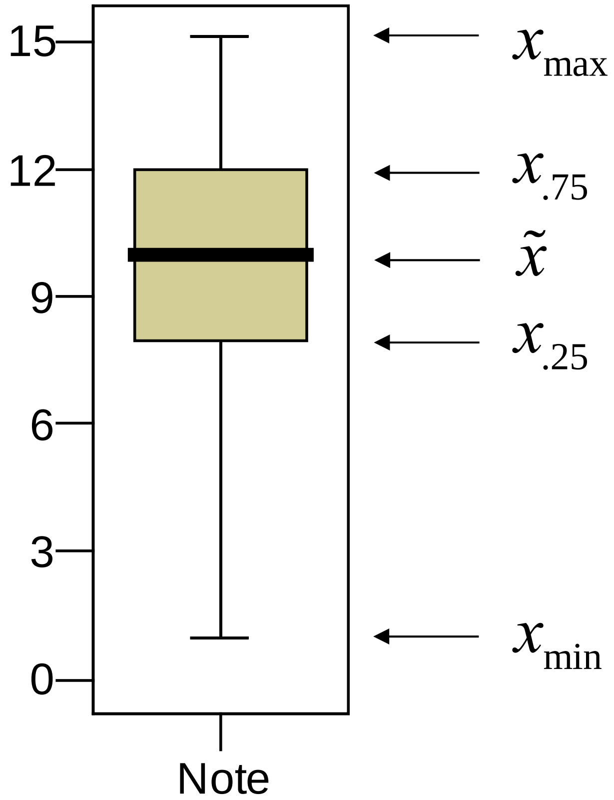
Bortz, S. 44 – 45

Ordinaldaten

Grafische Beschreibung: Box-Whisker-Plot

Mithilfe der **Fünf-Punkte-Zusammenfassung** (x_{min}, $x_{.25}$, x_{med}, $x_{.75}$, x_{max}) können Häufigkeitsdaten grafisch am **Boxplot** veranschaulicht werden.

Diese Variante ist problematisch, weil **Ausreißer** die Länge der Whisker erheblich vergrößern können



Ordinalskala

Häufigkeiten

Kennwerte

Grafische Darstellung

Ordinaldaten

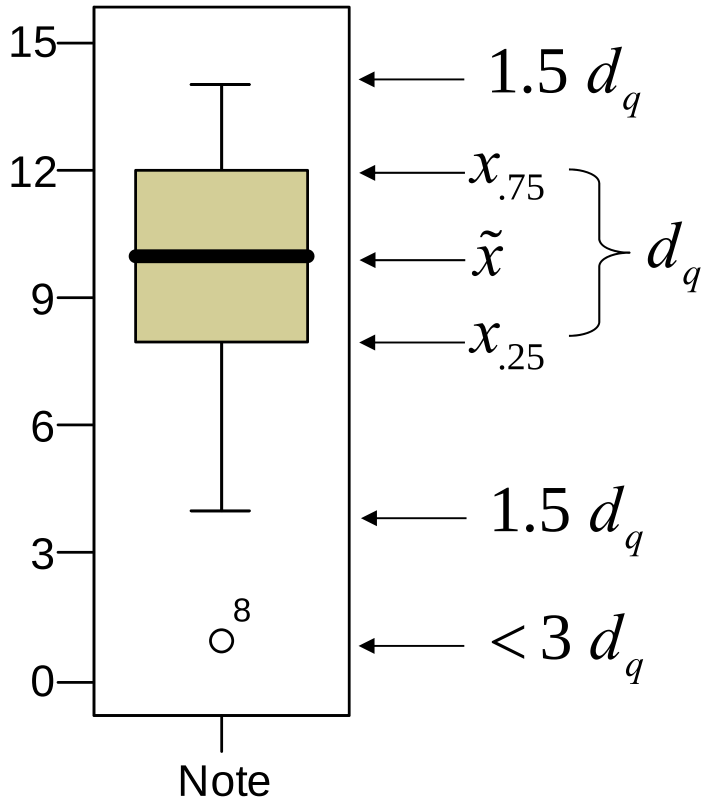
Grafische Beschreibung: Box-Whisker-Plot

Eine zweite, häufiger verwendete Variante des Boxplots verwendet den 1.5fachen **Interquartils-abstand** d_q für die Länge der Whisker.

Box und Whisker enden **am letzten Datenpunkt** innerhalb ihrer Reichweite

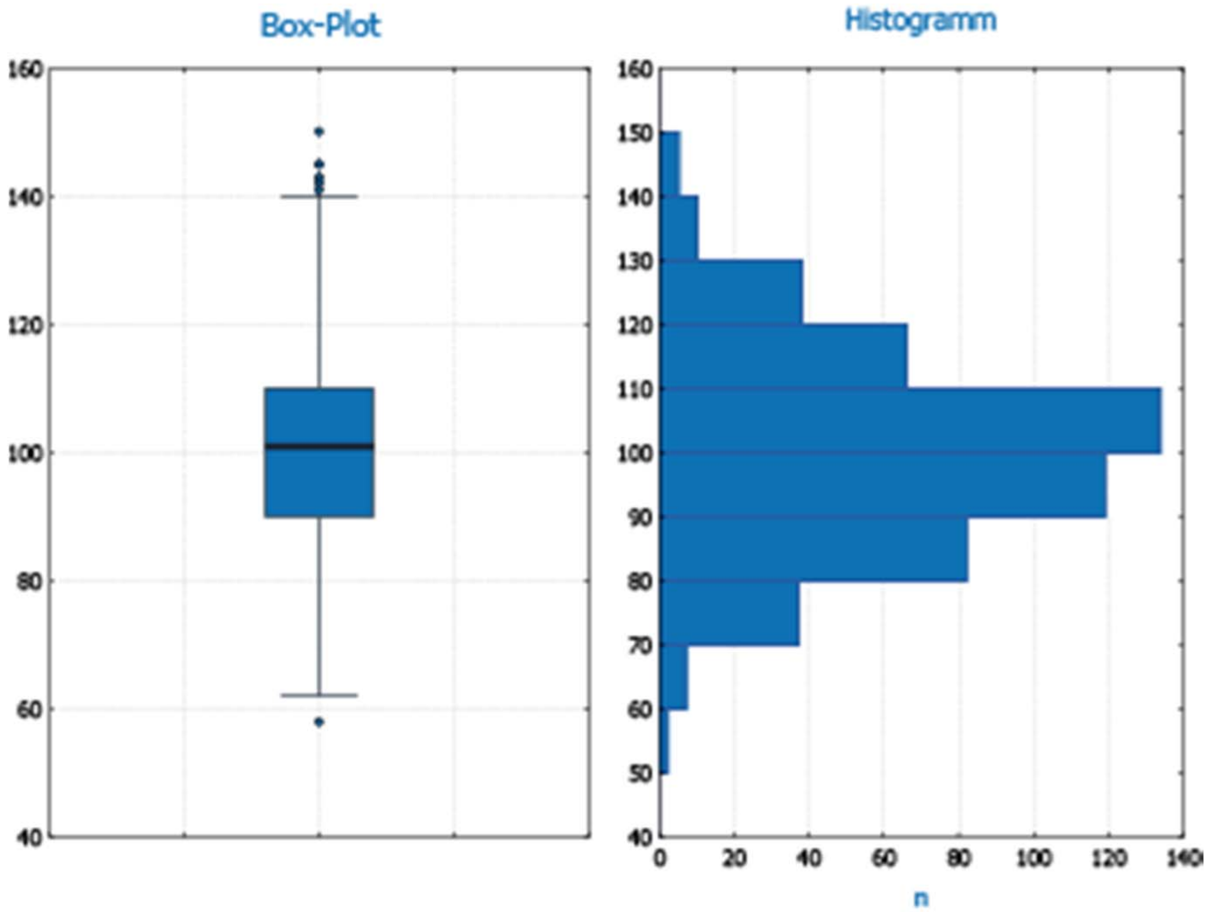
Datenpunkte außerhalb der Whisker werden explizit eingetragen (o).

Ausreißer $>3d_q$ werden mit Sternchen (*) markiert.



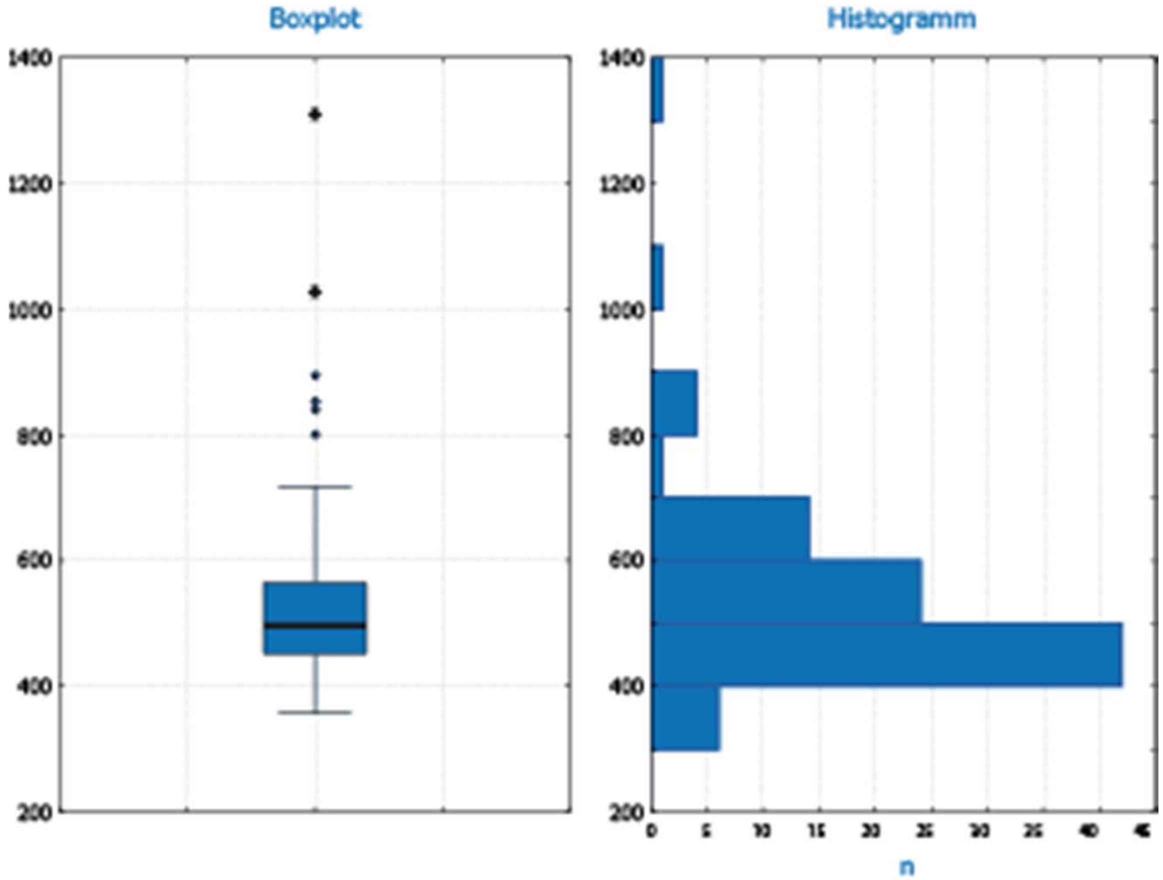
Ordinaldaten

Grafische Beschreibung: Box-Whisker-Plot



Ordinaldaten

Grafische Beschreibung: Box-Whisker-Plot



Bortz, S. 21 – 22

Intervallskala

Definition

- Es wird eine **Einheit** definiert
- Es existiert **kein natürlicher Nullpunkt**
- **Differenzen von Werten** können verglichen werden, nicht aber die Werte selbst
- Wird am häufigsten in empirischen psychologischen Untersuchungen angenommen
- Intervallskalierte Variablen können diskret oder stetig sein

Intervallskala

Beispiel

Attitudes Toward Housecleaning Scale von Ogletree, Worthen, Turner & Vickers (2006).

Ihre Aufgabe ist es, ihre Gefühle gegenüber jeder Aussage dahingehend zu kennzeichnen, ob sie (1) stark zustimmen, (2) etwas zustimmen, (3) weder zustimmen noch ablehnen, (4) etwas ablehnen oder (5) stark ablehnen. Bitte verdeutlichen Sie Ihre Meinung dadurch, dass sie entweder 1, 2, 3, 4 oder 5 auf dem Antwortblatt schwärzen.

- Einen Stapel dreckigen Geschirrs über Nacht im Spülbecken liegen zu lassen finde ich ekelhaft.
- Ich finde Staubwischen entspannend.
- Den Müll rauszubringen macht mir Spaß
- Frauen sollten die primäre Verantwortung für die Hausarbeit übernehmen.
- Eine unordentliche Wohnung zu haben macht mir nichts

Intervallskala

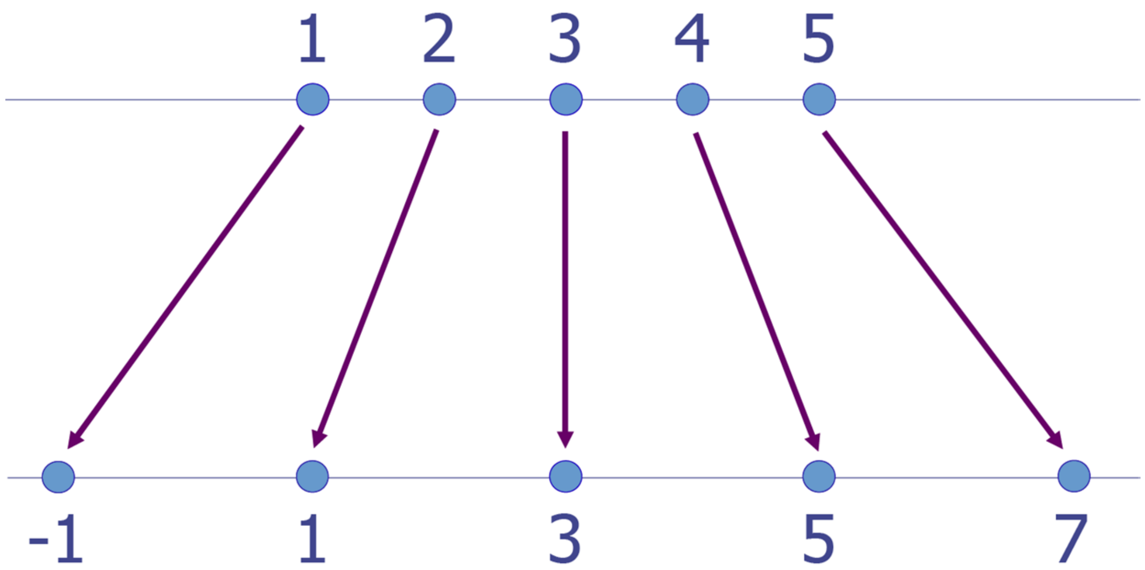
Zulässige Transformationen

- Zulässige Operationen sind **Äquivalenzrelationen**, d.h. **Gleich** und **Ungleich**
- **Zudem** erlaubt sind **qualitative Vergleichsrelationen**, d.h. **Größer** oder **Kleiner**
- Erlaubt sind weiterhin **quantitative Vergleichs-relationen**, die sich auf **Differenzen** beziehen
- Eine Aussage wie „*Der Unterschied zwischen A und B ist doppelt so groß wie zwischen A und C*" ist bei einer intervallskalierten Variable zulässig, nicht aber „*A ist doppelt so groß wie B*".

Intervallskala

Zulässige Transformationen

Zulässig sind alle **linearen Transformationen** (die Grundrechenarten), so dass die Verhältnisse zwischen Differenzen erhalten bleiben.



Intervallskala

Kreuztabellen

Grafische Darstellung I

Intervallskala

Zulässige Transformationen

Person	Distanz zwischen Personen	Skala 1	Skala 2	Skala 3	Skala 4
A		0	0.0	-1.0	70
	2				
B		2	1.0	0.0	90
	1				
C		3	1.5	0.5	100
	2				
D		5	2.5	1.5	120
	1				
E		6	3.0	2.0	130

Die Aussage „*Person E ist doppelt so gut wie Person C*", ausgehend von Skala 1, **gilt nicht** für Skala 3 und 4.

Intervallskala

Kreuztabellen

Grafische Darstellung I

Intervallskala

Zulässige Transformationen

Person	Distanz zwischen Personen	Skala 1	Skala 2	Skala 3	Skala 4
A		0	0.0	-1.0	70
	2				
B		2	1.0	0.0	90
	1				
C		3	1.5	0.5	100
	2				
D		5	2.5	1.5	120
	1				
E		6	3.0	2.0	130

Wohl aber **gilt immer**: *„Der Unterschied zwischen A und B ist doppelt so groß wie zwischen B und C"*

Intervallskala

Kritische Betrachtung

- Die bekanntesten und am meisten verbreiteten statistischen Verfahren setzen eine Intervallskala voraus
- Der Umgang mit niedrigeren Skalenniveaus ist mathematisch oftmals weitaus komplexer
- Die ungeprüfte Annahme der Intervallskala in psychologischen Untersuchungen ist oft problematisch
- Beispiele: Schulnoten, IQ-Skalen, Likert Skalen, Becks Depressionsskala (BDI)

0 – 13:	Keine bis minimale Depression
14 – 19:	Milde Depression
20 – 28:	Moderate Depression
29 – 63:	Schwere Depression

Intervalldaten

Numerische Beschreibung: Kreuztabellen

- **Problem**: Intervallskalierte Variablen können u.U. beliebige Ausprägungen besitzen, die sich nicht mehr sinnvoll in einer Tabelle darstellen lassen

- **Beispiele**: Körpergrößen, Serotoninspiegel, Reaktionszeit

- **Lösung:** Es muss eine **Aggregation** vieler Ausprägungen in wenige Kategorien (oder „Klassen") stattfinden

- Bei der Klassenbildung für eine Variable X findet im Prinzip nichts anderes als eine **Transformation von X** in eine neue Variable Y statt, und zwar gemäß

$$Y = \begin{cases} y_1 : X = \{\ldots\} \\ y_2 : X = \{\ldots\} \\ \ldots \\ y_k : X = \{\ldots\} \end{cases}$$

Intervalldaten

Numerische Beschreibung: Klassenbildung

- Die Messwertklassen dürfen sich nicht überschneiden, sie sind also wechselseitig ausschließend.
- Die untere und obere Klassengrenze UG_j und OG_j gehören zur Klasse c_j, die obere Grenze der vorherigen Klasse OG_{j-1} jedoch nicht.

  $\boxed{c_j = [UG_j \ldots OG_j]}$ oder $\boxed{c_j = (OG_{j-1} \ldots UG_{j+1}]}$

- Alle Klassen haben im Normalfall dieselbe Breite.
- Die Anzahl der Klassen ist zunächst frei wählbar. Es ist aber zu beachten:
  1. Es sollte möglichst wenige leere Klassen geben
  2. Es sollten keine in den Daten enthaltenen wichtigen Informationen „herausggregiert" werden (z.B. mehrere Modalwerte)

Intervalldaten

Numerische Beschreibung: Klassenbildung

- Zur Bestimmung der Anzahl von Klassen gibt es verschiedene Formeln. Als Faustregeln gelten:

Anzahl der Ausprägungen k	Klassenzahl c
5 bis 50	5 bis 8
50 bis 100	6 bis 10
100 bis 250	7 bis 12
>250	8 bis 25

- Eine einfache Formel, die oft zu einer sinnvollen Klassenanzahl c führt, lautet

$$\boxed{c = \left\lceil \log_2(n) + 1 \right\rceil} \quad \text{mit } \lceil \ \rceil = \text{Aufrundung}$$

- Statt der Beobachtungen n kann auch die Anzahl der Realisationen k verwendet werden.

Intervalldaten

Numerische Beschreibung: Klassenbildung

- Die **Klassenbreite** d bei einer gewünschten Anzahl von c gleich breiten Klassen wird berechnet als

$$\boxed{d = \frac{\max(X) - \min(X) + u}{c}}$$

mit u = Einheit der Skala (z.B. 1 oder 0.5)

- Hier ist X die ursprüngliche intervallskalierte Variable

- Bei der Berechnung der Klassenbreite muss auf **Ausreißer** in der Variablen X geachtet werden, da solche die Klassenbreite erheblich verzerren können.

Intervalldaten

Numerische Beschreibung: Klassenbildung

25 Abiturienten erreichen in ihrer Abschlussarbeit folgende Punktzahlen:

```
(11, 15, 8, 13, 8, 11, 14, 11, 11, 14, 13, 11,
2, 9, 10, 10, 14, 7, 7, 12, 12, 8, 6, 11, 13)
```

- Unter der Annahme, dass die Notenskala von 1 bis 15 reicht, ergibt sich diese Häufigkeitstabelle bei 5 Klassen:

Note	h(x)	f(x)	F(x)
1 – 3	1	0.04	0.04
4 – 6	2	0.08	0.12
7 – 9	6	0.28	0.40
10 – 12	10	0.44	0.84
13 – 15	7	0.16	1.00

Intervallskala

Kreuztabellen

Grafische
Darstellung I

Intervalldaten

Numerische Beschreibung: Klassenbildung

25 Ratten erreichen in einem Experiment folgende Reaktionszeiten:

```
(11.23, 15.1, 8.4576, 13.3, 8.955, 11.0, 14.443, 11.63,
11.39, 14.771, 13.115, 11.32, 2.5, 9.814, 10.03, 10.99,
14.3, 7.523, 7.49, 12.1496, 12.88, 8.0, 6.748, 11.1, 13.0)
```

- Schreibweise der Klassengrenzen in der Tabelle?

Note	h(x)	f(x)	F(x)
1 – 3	1	0.04	0.04
3 – 6	2	0.08	0.12
6 – 9	6	0.28	0.40
9 – 12	10	0.44	0.84
12 – 15	7	0.16	1.00

- Es galt per Konvention: **Die obere Grenze gehört zur Klasse, die untere nicht** (außer bei erster Kategorie).

Intervalldaten

Numerische Beschreibung: Klassenbildung

- Bei **diskreten Daten** werden die Klassengrenzen nach Möglichkeit **nicht-überlappend** angegeben.
- Die Klassenbreite ist dann
  mit u = Einheit der Skala (z.B. 1 oder 0.5)

$$\boxed{d = OG - UG + u}$$

- Bei **kontinuierlichen Daten** werden die Klassen-grenzen **überlappend** angegeben, wobei per Konvention die obere Grenze zur Klasse gehört, die untere aber nicht (mit Ausnahme der ersten Klasse).
- Die Klassenbreite ist dann

$$\boxed{d = OG - UG}$$

Relevante Excel Funktionen

- Kennwerte
  - MIN(), MAX()
  - MEDIAN()
  - QUANTIL().INKL(), QUARTILE.INKL(), QUANTILSRANG.INKL()
- Klassenbildung
  - LOG()
  - AUFRUNDEN()